# Supporting Information

# for

# Analysis of Protein Dynamics Simulations

# by a Stochastic Point Process Approach

Bertil Halle and Filip Persson

Biophysical Chemistry, Lund University, POB 124, SE-22100 Lund, Sweden

# S1. CODE FOR RT HISTOGRAM

The following Matlab code computes the RT histogram $\boldsymbol{F}_{\mathrm{R}}$, given the occupancy vector $\boldsymbol{A}_0$ (including any vacancies) for a singly occupied site.

```
A = A0(A0~=0);

B = any(diff(A,1,1)~=0,2);

B = [true; B(:); true];

VR = diff(find(B));

if A0(length(A0))~=0, VR(length(VR)) = []; end

if A0(1)~=0, VR(1) = []; end

FR = zeros(max(VR),1);

VRuni = unique(VR);

FR(VRuni) = histc(VR,VRuni);
```

# S2. CODE FOR RT AND ST STATISTICS

The following Matlab code computes the quantities $Q_{\mathrm{R}}(n)$, $Q_{\mathrm{S}}(n)$, $\tau_{\mathrm{R}}$ and $\tau_{\mathrm{S}}$, given the RT vector $\boldsymbol{V}_{\mathrm{R}}$, the RT histogram $\boldsymbol{F}_{\mathrm{R}}$ and the sampling resolution $\Delta\tau$. In addition, the uncorrelated statistical uncertainties in these quantities are computed.

```
Nvec = [0; (1:length(FR))'];

tau = dtau*Nvec;

NR = sum(FR);

NF = sum(VR);

VRm = NF/NR;

VR2m = mean(VR.^2);

VR3m = mean(VR.^3);

VR4m = mean(VR.^4);

tauR = dtau*VRm;

tauS = (dtau/2)*VR2m/VRm;

FRcumsum1 = flipud(cumsum(flipud(FR)));

QR = [1; (FRcumsum1 - FR)/NR];

FRcumsum2 = flipud(cumsum(flipud(FRcumsum1)));

QS = [1; (FRcumsum2 - FRcumsum1)/NF];

sigtauR = dtau*sqrt((VR2m - VRm^2)/(NR-1));

sigtauS = dtau*sqrt((VRm^2*VR4m + VR2m^3 - 2*VRm*VR2m*VR3m)/NR)/(2*VRm^2);

sigQR = sqrt(QR.*(1-QR)/(NR-1));

NQR = (2*Nvec+1).*QR;

sigQS = (dtau/tauR)*sqrt((cumsum(NQR) - NQR - (cumsum(QR) - QR).^2)/NR);
```

# S3. MULTIPLE OCCUPANCY

Here we indicate how the algorithm for constructing the RT vector can be generalized to multiply occupied hydration sites.

For a site that contains $\nu$ water molecules at all times, we replace the occupation vector by an occupation matrix $\mathbf{A}$ of size $N_{\mathrm{F}} \times \nu$. To analyze the overall RT and ST statistics for the entire site, we need not keep track of the individual subsites or of any water interchange among them. Exchange events are identified by taking $\nu \times \nu$ differences of $w$ indices between successive frames. The site-based RTs obtained in this way are then multiplied by $\nu$ to obtain molecule-based RTs averaged over the $\nu$ subsites.

If the occupancy fluctuates in time, the $\mathbf{A}_0$ matrix is dimensioned according to the maximum occupancy $\nu_{\max}$ and a negative integer is entered instead of the molecule index $w$ whenever one or more subsites are vacant (for example, $-1$ and $-2$ if two subsites are vacant in the same frame). The occupancy in a given frame $k$ then equals the number of positive elements in the corresponding row of $\mathbf{A}_0$. Rather than removing all vacancies, as we did for the singly occupied site, we can treat the negative $w$ indices as additional water molecules. The first-frame index must now be accompanied by a label that specifies whether it refers to a vacancy or not, so that the corresponding vacancy RTs can be discarded.

# S4. BINNING ERROR

Here we provide more details about the analysis of the binning error in the RT histogram $F_{\mathrm{R}}(n)$ and in the RT and ST statistics.

Figure 1 shows a small part of the analyzed trajectory. On the upper continuous time line, exchange events are indicated by tick marks and the different resident water molecules are labeled by letters. On the lower discrete time line, the observation time points, that is, the frames saved for analysis, are indicated by dots separated by the time interval $\Delta\tau$, which is the resolution. For each frame, the identity of the observed resident water molecule is indicated.

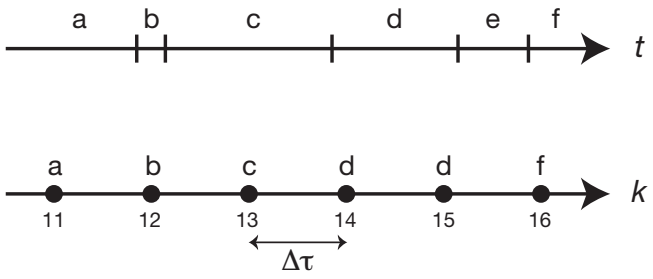


Figure S1: Top: continuous time line with 6 resident water molecules (labelled $a$ to $f$) and 5 exchange events (tick marks). Bottom: discrete time line indicating the water molecules observed at resolution $\Delta\tau$.

The distribution of RTs on the upper time line, that is, the separation of adjacent tick marks, is described by the continuous probability density $\psi_{\rm R}(\tau)$, such that $\psi_{\rm R}(\tau)\,{\rm d}\tau$ is the fraction of all RTs that are within ${\rm d}\tau$ of $\tau$ in length. The total number of RTs in the trajectory is denoted by $N_{\rm R}^0$.

From the lower time line, we obtain the observed RT histogram $F_{\rm R}(n)$, which gives the number of times that we observe in the trajectory the same resident water molecule in precisely $n$ contiguous frames. The total number of such discretized RTs is denoted by $N_{\rm R}$. Clearly,

$$N_{\rm R} \;=\; \sum_{n=1}^{\infty} F_{\rm R}(n) \tag{S1}$$

In Fig. S1, $N_{\rm R}^0 = 6$ but $N_{\rm R} = 5$ since water molecule $e$ escapes detection.

From the example in Fig. S1, it is seen that water molecules $b$ and $c$ are both assigned a discrete RT of $n = 1$ although the true continuous RTs are only a fraction of $\Delta\tau$ or nearly $2\,\Delta\tau$, respectively. In general, if the discrete RT is $n$, we know that the continuous RT must lie in the interval

$$(n-1)\,\Delta\tau \;<\; \tau \;\leq\; (n+1)\,\Delta\tau \tag{S2}$$

But not all continuous RTs in this interval contribute to $F_{\rm R}(n)$. The mapping from the continuous probability density $\psi_{\rm R}(\tau)$ to the histogram $F_{\rm R}(n)$ takes the form

$$F_{\rm R}(n) \;=\; N_{\rm R}^0 \int_{(n-1)\Delta\tau}^{(n+1)\Delta\tau} {\rm d}\tau\; \psi_{\rm R}(\tau)\, p_n(\tau) \quad , \quad n \geq 1 \tag{S3}$$

where $N_{\rm R}^0\,\psi_{\rm r}(\tau)\,{\rm d}\tau$ is the number of continuous RTs with length within ${\rm d}\tau$ of $\tau$ and $p_n(\tau)$ is the probability that a randomly placed time interval of length $\tau$ covers exactly $n$ frames. Clearly,

$$p_n(\tau) \;=\; \begin{cases} \dfrac{\tau-(n-1)\,\Delta\tau}{\Delta\tau} & , \quad \text{for} \qquad (n-1)\,\Delta\tau < \tau \le n\,\Delta\tau \\ \dfrac{(n+1)\,\Delta\tau-\tau}{\Delta\tau} & , \quad \text{for} \qquad n\,\Delta\tau < \tau \le (n+1)\,\Delta\tau \end{cases} \tag{S4}$$

Thus,

$$F_{\rm R}(n) = \frac{N_{\rm R}^0}{\Delta\tau}\left\{\int_{(n-1)\Delta\tau}^{n\Delta\tau}{\rm d}\tau\;\psi_{\rm R}(\tau)\,[\tau-(n-1)\,\Delta\tau] \;+\; \int_{n\Delta\tau}^{(n+1)\Delta\tau}{\rm d}\tau\;\psi_{\rm R}(\tau)\,[(n+1)\,\Delta\tau-\tau]\right\} \tag{S5}$$

For $\tau$ in the interval specified in eq S2, we can Taylor expand $\psi_{\rm R}(\tau)$ around the central frame at $\tau = n\Delta\tau$ as

$$\psi_{\rm R}(\tau) \;=\; \psi_{\rm R}(n\Delta\tau) + (\tau-n\Delta\tau)\,\frac{\partial\psi_{\rm R}}{\partial\tau}\Big|_{\tau=n\Delta\tau} + \frac{(\tau-n\Delta\tau)^2}{2}\,\frac{\partial^2\psi_{\rm R}}{\partial\tau^2}\Big|_{\tau=n\Delta\tau} + \cdots \tag{S6}$$

Inserting this expansion into eq S5 and carrying out the integrals, we find that the linear term in eq S6 does not contribute and

$$F_{\rm R}(n) \;=\; N_{\rm R}^0\,\Delta\tau\,\psi_{\rm R}(n\Delta\tau)\left[1 \;+\; \mathcal{O}\!\left(\frac{\Delta\tau}{\tau_{\rm R}}\right)^2\right] \quad , \quad n \ge 1 \tag{S7}$$

where, in the quadratic term, we have assumed that $\psi_{\rm R}''(\tau) = \psi_{\rm R}(\tau)/\tau_{\rm R}^2$, as for a Poisson process.

The discretized correlation functions $Q_{\rm X}(n)$ are most conveniently expressed, not in terms of the observed histogram $F_{\rm R}(n)$, but in terms of an 'ideal' histogram $F_{\rm R}^0(n)$

defined, for $n \geq 1$, as

$$F_{\mathrm{R}}^{0}(n) \;=\; N_{\mathrm{R}}^{0} \int_{(n-\frac{1}{2})\Delta\tau}^{(n+\frac{1}{2})\Delta\tau} \mathrm{d}\tau \, \psi_{\mathrm{R}}(\tau) \quad , \quad n \geq 1 \tag{S8}$$

and, for $n = 0$, as

$$F_{\mathrm{R}}^{0}(0) \;=\; N_{\mathrm{R}}^{0} \int_{0}^{\frac{\Delta\tau}{2}} \mathrm{d}\tau \, \psi_{\mathrm{R}}(\tau) \tag{S9}$$

Inserting the Taylor expansion from eq S6 and performing the integrals in eqs S8 and S9, we find

$$F_{\mathrm{R}}^{0}(n) \;=\; N_{\mathrm{R}}^{0} \, \Delta\tau \, \psi_{\mathrm{R}}(n\Delta\tau) \left[ 1 \;+\; \mathcal{O}\!\left( \frac{\Delta\tau}{\tau_{\mathrm{R}}} \right)^{2} \right] \quad , \quad n \geq 1 \tag{S10}$$

and

$$F_{\mathrm{R}}^{0}(0) \;=\; N_{\mathrm{R}}^{0} \, \Delta\tau \, \psi_{\mathrm{R}}(0) \left[ 1 \;+\; \mathcal{O}\!\left( \frac{\Delta\tau}{\tau_{\mathrm{R}}} \right) \right] \tag{S11}$$

Comparison of eqs S7 and S10 shows that

$$F_{\mathrm{R}}^{0}(n) \;=\; F_{\mathrm{R}}(n) \left[ 1 \;+\; \mathcal{O}\!\left( \frac{\Delta\tau}{\tau_{\mathrm{R}}} \right)^{2} \right] \quad , \quad n \geq 1 \tag{S12}$$

Thus, for $n \geq 1$, $F_{\mathrm{R}}^{0}(n)$ and $F_{\mathrm{R}}(n)$ are equal to first order in $\Delta\tau/\tau_{\mathrm{R}}$, that is, if we neglect terms of order $(\Delta\tau/\tau_{\mathrm{R}})^{2}$.

To establish the relation between $N_{\mathrm{R}}$ and $N_{\mathrm{R}}^{0}$, we first note that $F_{\mathrm{R}}(0)$ is undefined, since we cannot observe less than one frame. It follows immediately from eqs S8 and S9 that

$$\sum_{n=0}^{\infty} F_{\mathrm{R}}^{0}(n) \;=\; N_{\mathrm{R}}^{0} \tag{S13}$$

showing that no RTs are lost in this mapping. Thus,

$$N_{\rm R}^0 \;=\; F_{\rm R}^0(0) \;+\; \sum_{n=1}^{\infty} F_{\rm R}^0(n) \;\approx\; F_{\rm R}^0(0) \;+\; \sum_{n=1}^{\infty} F_{\rm R}(n) \;=\; F_{\rm R}^0(0) \;+\; N_{\rm R} \tag{S14}$$

where, according to eq S12, the second equality holds to first order in $\Delta\tau/\tau_{\rm R}$ and the last equality follows from eq S1. With eq S11, we then obtain to first order in $\Delta\tau/\tau_{\rm R}$

$$N_{\rm R} \;=\; N_{\rm R}^0\,[1 - \Delta\tau\,\psi_{\rm R}(0)] \tag{S15}$$

Since $\psi_{\rm R}(0)$ is of order $1/\tau_{\rm R}$ (at least for a Poisson process), we see that $N_{\rm R}$ and $N_{\rm R}^0$ differ to first order in $\Delta\tau/\tau_{\rm R}$, whereas $F_{\rm R}(n)$ and $F_{\rm R}^0(n)$ for $n \geq 1$ differ only to second order in $\Delta\tau/\tau_{\rm R}$.

We now consider the binning error in the RT and ST statistics. The trajectory length $T = N_{\rm F}\,\Delta\tau$ can be expressed either in terms of the observed number $N_{\rm R}$ of discrete RTs or in terms of the number $N_{\rm R}^0$ of continuous RTs:

$$T \;=\; N_{\rm F}\,\Delta\tau \;=\; N_{\rm R}\,\tau_{\rm R} \;=\; N_{\rm R}^0\,\tau_{\rm R}^0 \tag{S16}$$

where $N_{\rm F}$ is the number of frames or bins in the trajectory. The frames are labeled $k = 1, 2, \ldots, N_{\rm F}$. There are then $N_{\rm F} - 1$ intervals $\Delta\tau$. But the bins used in connection with $F_{\rm R}^0(n)$ include the $n = 0$ bin, so they also number $N_{\rm F}$.

The mean of the continuous RTs is

$$\tau_{\rm R}^0 \;=\; \Delta\tau\,\frac{N_{\rm F}}{N_{\rm R}^0} \tag{S17}$$

But the mean RT that we compute from the observed RTs is

$$\tau_{\mathrm{R}} \;=\; \Delta\tau\,\frac{N_{\mathrm{F}}}{N_{\mathrm{R}}} \tag{S18}$$

Consequently

$$\tau_{\mathrm{R}} \;=\; \tau_{\mathrm{R}}^0\,\frac{N_{\mathrm{R}}^0}{N_{\mathrm{R}}} \;=\; \frac{\tau_{\mathrm{R}}^0}{1-\Delta\tau\,\psi_{\mathrm{R}}(0)} \tag{S19}$$

where eq S15 was used in the last step. This is a systematic error: $\tau_{\mathrm{R}}$ is slightly longer than the true mean RT $\tau_{\mathrm{R}}^0$ because RTs shorter than $\Delta\tau$ may escape detection at a resolution of $\Delta\tau$. The mean RT $\tau_{\mathrm{R}}$ computed from eq S18 is therefore accurate only to zeroth order in $\Delta\tau/\tau_{\mathrm{R}}$, that is, it is too long by a relative amount of order $\Delta\tau/\tau_{\mathrm{R}}$.

The mean continuous ST can be computed from the identity

$$\tau_{\mathrm{S}}^0 \;=\; \frac{\langle\tau^2\rangle_{\mathrm{R}}^0}{2\,\tau_{\mathrm{R}}^0} \;=\; \frac{N_{\mathrm{R}}^0}{2\,N_{\mathrm{F}}\,\Delta\tau}\int_0^\infty \mathrm{d}\tau\,\tau^2\,\psi_{\mathrm{R}}(\tau) \tag{S20}$$

where eq S17 was used in the second step. By substituting the Taylor expansion of $\psi_{\mathrm{R}}(\tau)$ from eq S6 into eq S20 and carrying out the integrals, we find to first order in $\Delta\tau/\tau_{\mathrm{R}}$, that is, by neglecting terms of order $(\Delta\tau)^2/(\tau_{\mathrm{R}}\,\tau_{\mathrm{S}})$ or higher,

$$\tau_{\mathrm{S}} \;=\; \frac{\Delta\tau}{2\,N_{\mathrm{F}}}\sum_{n=1}^{\infty} n^2\,F_{\mathrm{R}}(n) \tag{S21}$$

Because this expression does not involve $N_{\mathrm{R}}$, the mean ST $\tau_{\mathrm{S}}$ computed from eq S21 is, like $F_{\mathrm{R}}(n)$ for $n \geq 1$, accurate to first order in $\Delta\tau/\tau_{\mathrm{R}}$. Discretization introduces a

second-order error in both $\tau_{\rm R}$ and $\tau_{\rm S}$. However, the resolution error is of first order in $\tau_{\rm R}$ but only of second order in $\tau_{\rm S}$. The resolution error (loss of short RTs) impacts less on $\tau_{\rm S}$ than on $\tau_{\rm R}$ because short RTs contribute less to $\tau_{\rm S}$.

The continuum residence correlation function (RCF) $Q^0_{\rm R}(\tau)$ is defined as

$$Q^0_{\rm R}(\tau) \;=\; \int_\tau^\infty {\rm d}\tau'\, \psi_{\rm R}(\tau') \tag{S22}$$

Thus, $Q^0_{\rm R}(\tau)$ is the fraction of continuous RTs that are longer than $\tau$. The discrete RCF $Q^0_{\rm R}(n)$ is defined as the fraction of bin-based RTs (of length $n\,\Delta\tau$, including the possibility of $n = 0$) that are longer than $n\,\Delta\tau$. Equation S22 yields to first order

$$Q^0_{\rm R}(n) \;=\; \frac{1}{N^0_{\rm R}} \sum_{p=n+1}^{\infty} F_{\rm R}(p) \tag{S23}$$

where we have used eq S12 to replace $F^0_{\rm R}(n)$ by $F_{\rm R}(n)$, which is also accurate to first order. But the RCF that we compute is

$$Q_{\rm R}(n) \;=\; \frac{1}{N_{\rm R}} \sum_{p=n+1}^{\infty} F_{\rm R}(p) \tag{S24}$$

which has $N_{\rm R}$ rather than $N^0_{\rm R}$. Therefore, in view of eq S15,

$$Q_{\rm R}(n) \;=\; \frac{Q^0_{\rm R}(n)}{1 - \Delta\tau\,\psi_{\rm R}(0)} \tag{S25}$$

showing that $Q_{\rm R}(n)$ is only accurate to zeroth order in $\Delta\tau/\tau_{\rm R}$.

The continuum survival correlation function (SCF) $Q^0_{\rm S}(\tau)$ is related to the RCF via the identity

$$\frac{{\rm d}\,Q^0_{\rm S}(\tau)}{{\rm d}\tau} \;=\; -\,\frac{Q^0_{\rm R}(\tau)}{\tau^0_{\rm R}} \tag{S26}$$

which integrates to

$$Q_{\mathrm{S}}^{0}(\tau) \;=\; \frac{1}{\tau_{\mathrm{R}}^{0}} \int_{\tau}^{\infty} \mathrm{d}\tau' \, Q_{\mathrm{R}}^{0}(\tau') \tag{S27}$$

The discrete version of this expression, accurate to first order, is

$$Q_{\mathrm{S}}^{0}(n) \;=\; \frac{N_{\mathrm{R}}^{0}}{N_{\mathrm{F}}} \sum_{p=n}^{\infty} Q_{\mathrm{R}}^{0}(p) \tag{S28}$$

where we have used eq S17. If we now substitute $Q_{\mathrm{R}}^{0}(p)$ from eq S23, the unknown quantity $N_{\mathrm{R}}^{0}$ cancels out and we are left with

$$Q_{\mathrm{S}}^{0}(n) \;=\; \frac{1}{N_{\mathrm{F}}} \sum_{p=n}^{\infty} \sum_{q=p+1}^{\infty} F_{\mathrm{R}}(q) \tag{S29}$$

This expression, which only involves the known quantities $N_{\mathrm{F}}$ and $F_{\mathrm{R}}(q)$, is also the one used to compute the approximate $Q_{\mathrm{S}}(n)$ from the MD data. We thus see that $Q_{\mathrm{S}}(n)$ is accurate to first order, whereas $Q_{\mathrm{R}}(n)$ is only accurate to zeroth order.

# S5. STATISTICAL ERROR

Here we provide more details about the analysis of the statistical uncertainties, or standard deviations, $\sigma(\tau_{\rm X})$ and $\sigma(Q_{\rm X})$ in $\tau_{\rm X}$ and $Q_{\rm X}$ (with X = R or S), resulting from the finite length $T$ of the trajectory at a given sampling resolution $\Delta\tau$. Here, we assume that $\Delta\tau$ is so small that the binning error is negligible.

Let $\tau_{\rm X}(T)$ be the value of $\tau_{\rm X}$ that we compute from a trajectory of length $T$. This value generally deviates somewhat from the value $\tau_{\rm X}(T \to \infty)$ that we would compute from an infinitely long trajectory. The standard deviation $\sigma(\tau_{\rm X})$ is a measure of this deviation. If the system is ergodic, the infinite trajectory average is equal to the ensemble average,

$$\langle \tau_{\rm X} \rangle \;=\; \lim_{T\to\infty} \tau_{\rm X}(T) \tag{S30}$$

The ensemble is an effectively infinite set of finite trajectories. We can choose to work with either of two ensembles. In the F (frame) ensemble, all trajectories contain the same number $N_{\rm F}$ of frames and are therefore of the same length $T = N_{\rm F}\,\Delta\tau$. In the R (residence time) ensemble, all trajectories contain the same number $N_{\rm R}$ of residence time (RT) intervals. The trajectory length is then more conveniently expressed as $T = N_{\rm R}\,\tau_{\rm R}$. Because $\tau_{\rm R}$ fluctuates somewhat among the different trajectories, it follows that the trajectories in the $N_{\rm R}$ ensemble are not of precisely the same length.

Our primary data is a sequence of $N_{\mathrm{R}}$ RT intervals. The number $N_{\mathrm{R}}$ is thus a measure of the amount of information at our disposal. If we sample the trajectory more densely by decreasing $\Delta\tau$, we increase $N_{\mathrm{F}}$ but we do not gain more information since $N_{\mathrm{R}}$ is unchanged. The statistical error should therefore depend on $N_{\mathrm{R}}$, but not on $N_{\mathrm{F}}$. Specifically, if the RTs are mutually independent, we expect the statistical error to be proportional to $1/\sqrt{N_{\mathrm{R}}}$. The statistical error must therefore be evaluated in the $N_{\mathrm{R}}$ ensemble. In the following, we denote R ensemble averages by $\langle\cdots\rangle$, whereas F ensemble averages, in the few cases where they occur, are denoted by $\langle\cdots\rangle_{\mathrm{F}}$.

Let $n$ be the RT in units of $\Delta\tau$. The ensemble-averaged mean RT is then

$$\langle n\rangle \;=\; \sum_{n=1}^{\infty} n\, F_{\mathrm{R}}(n) \tag{S31}$$

where $F_{\mathrm{R}}(n)$ is the equilibrium RT probability distribution. Neither $F_{\mathrm{R}}(n)$ nor $\langle n\rangle$ can be obtained from a trajectory of finite length. From the trajectory, we obtain a chronological series of a finite number $N_{\mathrm{R}}$ of RTs. If the simulated system is in equilibrium, the series $\{n_\alpha\}_{\alpha=1\ldots N_{\mathrm{R}}}$ represents a stationary stochastic process. Stationarity implies, among other things, that the $n_\alpha$ are identically distributed, but not necessarily independent, random variables. Therefore, the mean $\langle n_\alpha\rangle \equiv \langle n\rangle$ does not depend on $\alpha$. An unbiased estimator of the mean RT $\langle n\rangle$ is provided by the

trajectory average

$$\overline{n}(N_{\mathrm{R}}) \;=\; \frac{1}{N_{\mathrm{R}}} \sum_{\alpha=1}^{N_{\mathrm{R}}} n_\alpha \tag{S32}$$

This estimator is unbiased because

$$\langle \overline{n}(N_{\mathrm{R}}) \rangle \;=\; \frac{1}{N_{\mathrm{R}}} \sum_{\alpha=1}^{N_{\mathrm{R}}} \langle n_\alpha \rangle \;=\; \langle n \rangle \tag{S33}$$

Note that $N_{\mathrm{R}}$ is not ensemble-averaged since it is a constant in the R ensemble. To simplify the notation, we will write $\overline{n}$ instead of $\overline{n}(N_{\mathrm{R}})$ in the following.

The trajectory average $\overline{n}$ may be regarded as a random variable with a probability density $\Psi(\overline{n})$. If $N_{\mathrm{R}}$ is large, the central limit theorem leads us to expect that $\Psi(\overline{n})$ is approximately Gaussian and hence fully characterized by the mean $\langle \overline{n} \rangle = \langle n \rangle$ and the variance

$$\sigma^2(\overline{n}) \;\equiv\; \left\langle [\overline{n} - \langle \overline{n} \rangle]^2 \right\rangle \;=\; \left\langle \overline{n}^2 \right\rangle \;-\; \langle \overline{n} \rangle^2 \tag{S34}$$

The desired measure of the statistical error in the estimator $\overline{n}$ of the mean RT is the standard deviation $\sigma(\overline{n})$. However, eq S34 involves ensemble averages, which we cannot obtain from a finite trajectory. We can only obtain an estimator $s^2(\overline{n})$ of the variance $\sigma^2(\overline{n})$. To be reliable, this should be an unbiased estimator.

Substituting $\overline{n}$ from eq S32 into eq S34, we obtain

$$\sigma^2(\overline{n}) \;=\; \frac{1}{N_{\mathrm{R}}^2} \sum_{\alpha=1}^{N_{\mathrm{R}}} \sum_{\beta=1}^{N_{\mathrm{R}}} C_{\alpha\beta} \tag{S35}$$

with the covariance $C_{\alpha\beta}$ of the random variables $n_\alpha$ and $n_\beta$ defined as

$$C_{\alpha\beta} \equiv \langle n_\alpha\, n_\beta \rangle - \langle n_\alpha \rangle \langle n_\beta \rangle = \langle n_\alpha\, n_\beta \rangle - \langle n \rangle^2 \tag{S36}$$

If the random variables $n_\alpha$ and $n_\beta$ are uncorrelated, we have $\langle n_\alpha\, n_\beta \rangle = \langle n_\alpha \rangle \langle n_\beta \rangle = \langle n \rangle^2$ and, therefore, $C_{\alpha\beta} = 0$. Moreover, $C_{\alpha\alpha} = \sigma_n^2$ is the variance of the random variable $n_\alpha$,

$$\sigma_n^2 \equiv \langle n_\alpha^2 \rangle - \langle n_\alpha \rangle^2 = \langle n^2 \rangle - \langle n \rangle^2 \tag{S37}$$

In general, $|C_{\alpha\beta}| \leq \sigma_n^2$.

It is convenient to split $\sigma^2(\overline{n})$ in two parts, corresponding to the diagonal and off-diagonal terms of the double sum in eq S35. Thus,

$$\sigma^2(\overline{n}) = \sigma_0^2(\overline{n}) + \Delta_{\rm c}(\overline{n}) \tag{S38}$$

$$\sigma_0^2(\overline{n}) = \frac{\sigma_n^2}{N_{\rm R}} \tag{S39}$$

$$\Delta_{\rm c}(\overline{n}) = \frac{1}{N_{\rm R}^2} \sum_{\alpha=1}^{N_{\rm R}} {\sum_{\beta=1}^{N_{\rm R}}}' C_{\alpha\beta} \tag{S40}$$

where the prime on the second sum signifies omission of diagonal ($\beta = \alpha$) terms. Whereas $\sigma^2(\overline{n})$ and $\sigma_0^2(\overline{n})$ are necessarily non-negative, the quantity $\Delta_{\rm c}$ may be negative.

If all the random variables $\{n_\alpha\}_{\alpha=1\ldots N_{\rm R}}$ are mutually uncorrelated, it follows from the foregoing that $\Delta_{\rm c}(\overline{n}) = 0$ and that $\sigma^2(\overline{n}) = \sigma_0^2(\overline{n})$ with

$$\sigma_0^2(\overline{n}) = \frac{\sigma_n^2}{N_{\rm R}} = \frac{\langle n^2 \rangle - \langle n \rangle^2}{N_{\rm R}} \tag{S41}$$

This is a well-known result: the standard error of the mean of $N$ independent measurements of a random variable equals the standard deviation of the random variable itself divided by the square root of the number of measurements.

To obtain an unbiased estimator $s_0^2(\overline{n})$ of $\sigma_0^2(\overline{n})$, we need an unbiased estimator $s_n^2$ of the variance $\sigma_n^2$. This is

$$s_n^2 \;=\; \frac{1}{(N_\mathrm{R}-1)} \sum_{\alpha=1}^{N_\mathrm{R}} \left(n_\alpha^2 \;-\; \overline{n}^2\right) \tag{S42}$$

with the denominator $N_\mathrm{R}-1$ rather than $N_\mathrm{R}$. In the absence of correlations, the unbiased estimator $s_0^2(\overline{n})$ of the variance $\sigma_0^2(\overline{n})$ of the unbiased estimator $\overline{n}$ of the mean RT $\langle n \rangle$ can, according to eqs S32, S41 and S42, be computed from

$$s_0^2(\overline{n}) \;=\; \frac{1}{(N_\mathrm{R}-1)} \left[ \left( \frac{1}{N_\mathrm{R}} \sum_{\alpha=1}^{N_\mathrm{R}} n_\alpha^2 \right) \;-\; \left( \frac{1}{N_\mathrm{R}} \sum_{\alpha=1}^{N_\mathrm{R}} n_\alpha \right)^2 \right] \tag{S43}$$

The desired statistical uncertainty in $\tau_\mathrm{R}/\Delta\tau$ is the square root of this.

We now return to the general case with correlations. If the set $\{n_\alpha\}_{\alpha=1\ldots N_\mathrm{R}}$ of random variables represents a stochastic process with the index $\alpha$ increasing monotonically with time, then the covariance $C_{\alpha\beta}$ in eq S36 can be regarded as a serial correlation function. If the process is stationary, the correlation function can only depend on the difference $\gamma = \beta - \alpha$. The double sum in eq S40 can therefore be reduced to a single sum as

$$\Delta_\mathrm{c}(\overline{n}) \;=\; \frac{2}{N_\mathrm{R}} \sum_{\gamma=1}^{N_\mathrm{R}-1} \left(1 - \frac{\gamma}{N_\mathrm{R}}\right) C_\gamma \tag{S44}$$

$$C_\gamma \equiv \langle n_\alpha\, n_{\alpha+\gamma}\rangle - \langle n_\alpha\rangle\,\langle n_{\alpha+\gamma}\rangle = \langle n_\alpha\, n_{\alpha+\gamma}\rangle - \langle n\rangle^2 \tag{S45}$$

where $\langle n_\alpha\, n_{\alpha+\gamma}\rangle$ is independent of $\alpha$. In our case, the 'time' point $\alpha$ is simply the serial number of a RT in the chronological RT list. Hence, we refer to $C_\gamma$ as a serial correlation function rather than a time correlation function. The serial correlation function $C_\gamma$ measures the correlation of two RTs that are separated by $\gamma - 1$ intervening RTs.

Without *a priori* knowledge about the serial correlation function $C_\gamma$, eq S44 is of limited utility. Probably the most robust and computationally efficient approach to this problem is the block renormalization (BR) method proposed by Flyvbjerg and Petersen.[1] In this method, we select the longest sequence of $N = 2^k$ ($k$ is an integer) RTs contained in the original RT series. For this pruned series, we compute the $\sigma(\overline{n})$ estimator from eq S43 (which is valid in the absence of correlations)

$$s = \left\{ \frac{1}{(N-1)} \left[ \left( \frac{1}{N} \sum_{\alpha=1}^{N} n_\alpha^2 \right) - \left( \frac{1}{N} \sum_{\alpha=1}^{N} n_\alpha \right)^2 \right] \right\}^{1/2} \tag{S46}$$

We then perform the first blocking transformation by averaging pairwise values

$$n'_\alpha = \frac{n_{2\alpha-1} + n_{2\alpha}}{2} \tag{S47}$$

so that $N' = N/2$. We then compute a new value for the estimator,

$$s' = \left\{ \frac{1}{(N'-1)} \left[ \left( \frac{1}{N'} \sum_{\alpha=1}^{N'} n_\alpha'^2 \right) - \left( \frac{1}{N'} \sum_{\alpha=1}^{N'} n'_\alpha \right)^2 \right] \right\}^{1/2} \tag{S48}$$

This process is repeated until $N' = 2$, yielding a series of $k$ values for the estimator. The relative standard deviation in these estimators is $[2\,(N'-1)]^{-1/2}$.[1] Whereas the mean $\overline{n}$ is invariant under the blocking transformation, the standard deviation estimator $s$ increases until a so-called fixed point is reached and then remains constant upon further transformations. Of course, such a plateau will only be observed if the correlation 'time' on which $C_\gamma$ decays is much smaller than $N_\mathrm{R}$.

Figure S2 shows estimates of $\sigma(\tau_\mathrm{R})$ after successive blocking transformations for site W113 in all states. Here, $N_\mathrm{R} = 37,499$, so that $k = 15$ ($2^{15} = 32,768$). Equation S43 yields $s_0(\tau_\mathrm{R}) \approx 2$ ns, but the BR estimate increases to $\sim 10$ ns without reaching a well-defined plateau. The source of the problem is evident from the RT vector in Fig. 2 of the main text: small and large RTs tend to cluster in alternating segments with a correlation 'time' of the same order of magnitude as $N_\mathrm{R}$. In Fig. S3, the BR method is applied to the RTs of site W113 in state M1 after omission of short RTs. Here the problem is the small number of RTs, $N_\mathrm{R} = 60$, only half of which can be used in the BR method ($2^5 = 32$). Because of this reduction in the number of RTs, the initial error in Fig. S3 is larger than the one computed from eq S43 with all 60 RTs. Although the uncertainty in $s(\tau_\mathrm{R})$ is large, the BR method indicates that correlations are unimportant, as is also suggested by the RT vector in Fig. 6 of the main text. The uncorrelated error $s_0(\tau_\mathrm{R})$ should then be a good estimate.

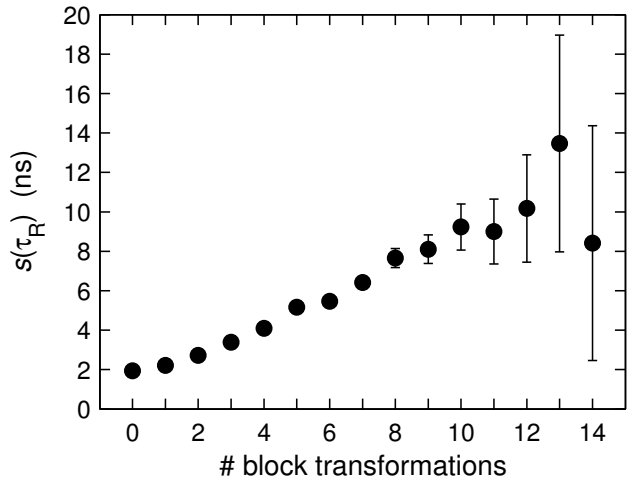


Figure S2: BR estimation of statistical error in $\tau_{\mathrm{R}}$ for site W113 in all states.

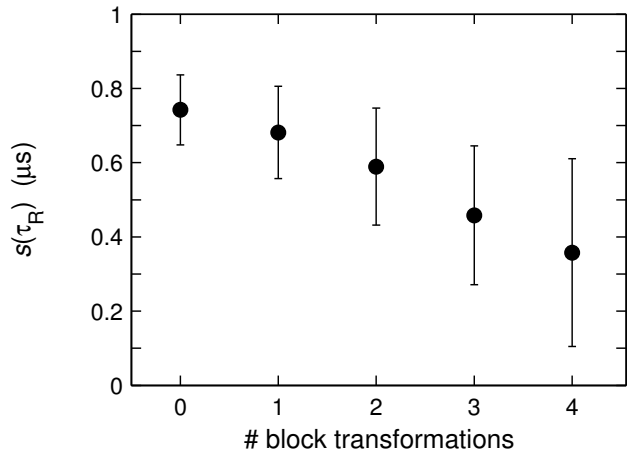


Figure S3: BR estimation of statistical error in $\tau_{\mathrm{R}}$ for site W113 in M1 state with exclusion of short ($< 50$ ns) RTs.

We now turn to the mean ST. Let $m_k$ be the number of remaining frames in the current RT, of length $n_\alpha$, beginning at, and including, an arbitrary frame $k$. This is the time to the next exchange event from an arbitrary starting time. From a finite-length trajectory, we obtain a chronological series of $N_\mathrm{F}$ STs. If the simulated system is in equilibrium, the series $\{m_k\}_{k=1\ldots N_\mathrm{F}}$ represents a stationary stochastic process and the trajectory average

$$\overline{m} \;=\; \frac{1}{N_\mathrm{F}} \sum_{k=1}^{N_\mathrm{F}} m_k \tag{S49}$$

is an unbiased estimator of the mean ST $\langle m \rangle_\mathrm{F}$ in the F ensemble.

Since the RT series $\{n_\alpha\}_{\alpha=1\ldots N_\mathrm{R}}$ contains all information about the underlying stochastic process, it must be possible to express $\overline{m}$ in terms of the RTs $n_\alpha$. Indeed, it is readily shown that (see eq 7 of the main text)

$$\overline{m} \;=\; \frac{\overline{n^2}}{2\,\overline{n}} \;=\; \frac{1}{2}\,\frac{\sum_{\alpha=1}^{N_\mathrm{R}} n_\alpha^2}{\sum_{\alpha=1}^{N_\mathrm{R}} n_\alpha} \tag{S50}$$

We seek the variance

$$\sigma^2(\overline{m}) \;\equiv\; \left\langle [\overline{m} - \langle \overline{m} \rangle]^2 \right\rangle \;=\; \left\langle \overline{m}^2 \right\rangle \;-\; \langle \overline{m} \rangle^2 \tag{S51}$$

We can choose to evaluate the ensemble averages either in the F ensemble or in the R ensemble. In the former case, we can substitute $\overline{m}$ from eq S49 and proceed as for $\tau_\mathrm{R}$. But we want an expression that involves the RTs $n_\alpha$, since the RT series

constitutes our data. We must therefore work with the R ensemble, where $N_{\mathrm{R}}$ is constant. Our starting point must therefore be eq S50, which expresses the random variable $\overline{m}$ as a nonlinear function of two other (correlated) random variables $\overline{n}$ and $\overline{n^2}$. The conventional procedure for computing the variance of a function of random variables is to Taylor expand the function. Expanding eq S50 to first order, we obtain

$$\overline{m} \;=\; \langle m \rangle \;-\; \frac{\langle n^2 \rangle}{2\,\langle n \rangle^2}\big[\,\overline{n} - \langle n \rangle\big] \;+\; \frac{1}{2\,\langle n \rangle}\big[\,\overline{n^2} - \big\langle n^2 \big\rangle\big] \;+\; \cdots \tag{S52}$$

where we have noted that $\langle \overline{n} \rangle = \langle n \rangle$ and $\left\langle \overline{n^2} \right\rangle = \langle n^2 \rangle$ in the R ensemble. This expansion shows that $\langle \overline{m} \rangle = \langle m \rangle$ to first order in the R ensemble, whereas this is an exact result in the F ensemble.

Forming the difference $\overline{m} - \langle m \rangle$ from eq S52, squaring it, and taking the ensemble average, we obtain the variance

$$\begin{aligned} \sigma^2(\overline{m}) \;=\; & \left(\frac{\langle n^2 \rangle}{2\,\langle n \rangle^2}\right)^2 \left\langle \Big[\,\overline{n} - \langle n \rangle\Big]^2 \right\rangle \;+\; \left(\frac{1}{2\,\langle n \rangle}\right)^2 \left\langle \Big[\,\overline{n^2} - \big\langle n^2 \big\rangle\Big]^2 \right\rangle \\ & -\; \frac{\langle n^2 \rangle}{2\,\langle n^3 \rangle} \left\langle \Big[\,\overline{n} - \langle n \rangle\Big] \Big[\,\overline{n^2} - \big\langle n^2 \big\rangle\Big] \right\rangle \end{aligned} \tag{S53}$$

Since we have neglected terms of higher than second order, this result is accurate only if the relative variances in $\overline{n}$ and $\overline{n^2}$, as well as their relative covariance, are small. In other words, the Taylor expansion method only predicts $\sigma(\overline{m})$ accurately if $\sigma(\overline{m})/\overline{m}$ is small. Furthermore, the estimator obtained by replacing all ensemble averages in eq S53 by trajectory averages is not unbiased.

Rather than pursuing the general case, we shall assume that the RTs $n_\alpha$ are mutually uncorrelated. This seems to be approximately true if we omit the short RTs. In any case, $\sigma_0^2(\overline{m})$ provides a useful lower bound on $\sigma^2(\overline{m})$. For uncorrelated RTs, we have

$$\begin{aligned}
\left\langle \left[ \overline{n} - \langle n \rangle \right]^2 \right\rangle &= \frac{\langle n^2 \rangle - \langle n \rangle^2}{N_\mathrm{R}} \\
\left\langle \left[ \overline{n^2} - \left\langle n^2 \right\rangle \right]^2 \right\rangle &= \frac{\langle n^4 \rangle - \langle n^2 \rangle^2}{N_\mathrm{R}} \\
\left\langle \left[ \overline{n} - \langle n \rangle \right] \left[ \overline{n^2} - \left\langle n^2 \right\rangle \right] \right\rangle &= \frac{\langle n^3 \rangle - \langle n \rangle \langle n^2 \rangle}{N_\mathrm{R}}
\end{aligned} \tag{S54}$$

which is inserted into eq S53 to yield

$$\sigma_0^2(\overline{m}) = \frac{1}{N_\mathrm{R}} \frac{\langle n \rangle^2 \langle n^4 \rangle + \langle n^2 \rangle^3 - 2 \langle n \rangle \langle n^2 \rangle \langle n^3 \rangle}{4 \langle n \rangle^4} \tag{S55}$$

As expected, the error $\sigma_0(\overline{m})$ decreases as the inverse square root of $N_\mathrm{R}$, just as for $\sigma_0(\overline{n})$. However, the estimator $s_0^2(\overline{m})$ obtained by replacing the ensemble averages in eq S55 by trajectory averages (and perhaps $N_\mathrm{R}$ by $N_\mathrm{R} - 1$) is not unbiased. Moreover, we cannot use the BR method to obtain $\sigma^2(\overline{m})$ (including the effect of correlations), because the trajectory average $\overline{m}$ in eq S50 is not invariant under blocking transformation. Nevertheless, we can hope that eq S55 yields a reasonable estimate if the relative error is small and if the correlations are insignificant.

If the $n_\alpha$ are not only mutually independent but also exponentially distributed,

they constitute a Poisson process and

$$\left\langle n^k \right\rangle \;=\; k!\,\langle n \rangle^k \tag{S56}$$

In this special case, eq S55 reduces to

$$\sigma_0^2(\overline{m}) \;=\; \frac{2\,\langle n \rangle^2}{N_{\mathrm{R}}} \tag{S57}$$

whereas the corresponding result in eq S41 for the variance of the mean RT becomes

$$\sigma_0^2(\overline{n}) \;=\; \frac{\langle n \rangle^2}{N_{\mathrm{R}}} \tag{S58}$$

Whereas $\tau_{\mathrm{S}} = \tau_{\mathrm{R}}$ (or $\overline{m} = \overline{n}$) for a Poisson process, the standard deviation is thus a factor $\sqrt{2}$ larger for $\tau_{\mathrm{S}}$.

The trajectory-based estimator of the residence correlation function (RCF) $Q_{\mathrm{R}}(n)$ is denoted by $\overline{h}(n)$. This may be expressed as a trajectory average

$$\overline{h}(n) \;=\; \frac{1}{N_{\mathrm{R}}} \sum_{\alpha=1}^{N_{\mathrm{R}}} h_\alpha(n) \tag{S59}$$

of the indicator function $h_\alpha(n)$, which equals 1 if $n_\alpha > n$ and 0 otherwise. The ensemble-averaged RCF is

$$\langle h(n) \rangle \;=\; \sum_{p=n+1}^{\infty} F_{\mathrm{R}}(p) \tag{S60}$$

where $F_{\mathrm{R}}(p)$ is the equilibrium RT probability distribution. The trajectory average $\overline{h}(n)$ is an unbiased estimator of the ensemble average $\langle h(n) \rangle$, because

$$\left\langle \overline{h}(n) \right\rangle \;=\; \frac{1}{N_{\mathrm{R}}} \sum_{\alpha=1}^{N_{\mathrm{R}}} \langle h_\alpha(n) \rangle \;=\; \langle h(n) \rangle \tag{S61}$$

where we have made use of the stationarity of the stochastic process $\{n_\alpha\}_{\alpha=1\ldots N_\mathrm{R}}$.

The error analysis for $\tau_\mathrm{R}$ carries over directly to the RCF if we replace $n_\alpha$ by $h_\alpha(n)$. To simplify the notation, we shall omit the argument $n$ in most of what follows. Because $h_\alpha$ can only take the values 0 or 1, the expressions can be further simplified by using the identity $h_\alpha^2 = h_\alpha$. The variance of $h_\alpha$ is thus given by

$$\sigma_h^2 \;\equiv\; \left\langle h_\alpha^2 \right\rangle - \langle h_\alpha \rangle^2 \;=\; \langle h \rangle \left[1 - \langle h \rangle\right] \tag{S62}$$

If the RTs are mutually uncorrelated, the variance of $\overline{h}$ is, in analogy with eq S41,

$$\sigma_0^2(\overline{h}) \;=\; \frac{\sigma_h^2}{N_\mathrm{R}} \;=\; \frac{\langle h \rangle \left[1 - \langle h \rangle\right]}{N_\mathrm{R}} \tag{S63}$$

and the unbiased estimator (see eq S43) for the standard deviation is

$$s_0(\overline{h}) \;=\; \left[\frac{\overline{h}\,(1-\overline{h})}{N_\mathrm{R}-1}\right]^{1/2} \tag{S64}$$

Since $\overline{h}(0) \equiv 1$, eq S64 predicts a vanishing error for $n = 0$, as expected.

The trajectory-based estimator of the survival correlation function (SCF) $Q_\mathrm{S}(n)$ is denoted by $\overline{g}(n)$. This may be expressed as a trajectory average

$$\overline{g}(n) \;=\; \frac{1}{N_\mathrm{F}} \sum_{k=1}^{N_\mathrm{F}} g_k(n) \tag{S65}$$

of the indicator function $g_k(n)$, which equals 1 if $m_k > n$ and 0 otherwise.

The ensemble-averaged SCF is

$$\langle g(n) \rangle \;=\; \sum_{p=n+1}^{\infty} F_\mathrm{S}(p) \tag{S66}$$

where $F_{\mathrm{S}}(p)$ is the equilibrium ST probability distribution. The trajectory average $\overline{g}(n)$ is an unbiased estimator of the ensemble average $\langle g(n)\rangle_{\mathrm{F}}$ in the F ensemble, because the stochastic process is stationary so that

$$\langle \overline{g}(n)\rangle_{\mathrm{F}} \;=\; \frac{1}{N_{\mathrm{F}}}\sum_{k=1}^{N_{\mathrm{F}}} \langle g_k(n)\rangle_{\mathrm{F}} \;=\; \langle g(n)\rangle_{\mathrm{F}} \tag{S67}$$

As for $\tau_{\mathrm{S}}$, we want to express the trajectory average in eq S65 in terms of the random variables $n_\alpha$, so we can take ensemble averages in the R ensemble. We therefore write

$$\overline{g}(n) \;=\; \sum_{\alpha=1}^{N_{\mathrm{R}}} p_\alpha\, g_\alpha(n) \tag{S68}$$

where $g_\alpha(n)$ is the mean of the indicator function $g_k(n)$ for the RT interval $\alpha$ and $p_\alpha = n_\alpha/N_{\mathrm{F}} = n_\alpha/(N_{\mathrm{R}}\,\overline{n})$ is the probability that the randomly chosen frame $k$ lies in this interval. Furthermore,

$$g_\alpha(n) \;=\; \frac{1}{n_\alpha}\sum_{i=1}^{n_\alpha} g_i(n) \;=\; \frac{1}{n_\alpha}\,(n_\alpha - n)\,h_\alpha(n) \tag{S69}$$

where the indicator function $h_\alpha(n)$, introduced in connection with the RCF, acts as a step function that ensures that $g_\alpha(n) = 0$ for $n > n_\alpha$. Combining eqs S68 and S69 and the expression for $p_\alpha$, we obtain

$$\overline{g}(n) \;=\; \frac{1}{N_{\mathrm{R}}\,\overline{n}}\sum_{\alpha=1}^{N_{\mathrm{R}}} (n_\alpha - n)\,h_\alpha(n) \tag{S70}$$

Note that, in eqs S69 and S70, $n$ is not a random variable, but merely an integer.

We seek the variance

$$\sigma^2(\overline{g}) \;\equiv\; \left\langle [\overline{g} - \langle \overline{g} \rangle]^2 \right\rangle \;=\; \left\langle \overline{g}^2 \right\rangle \;-\; \langle \overline{g} \rangle^2 \tag{S71}$$

Now eq S70 expresses the random variable $\overline{g}$ as a nonlinear function of $2\,N_{\rm R} + 1$ other (correlated) random variables, namely $n_\alpha$, $h_\alpha$ and $\overline{n}$. Taylor expanding $\overline{g}(n)$ in eq S70 to first order in the fluctuations of these variables around their ensemble averages, we can proceed along the same lines as in the error analysis of $\tau_{\rm S}$. However, this leads to a rather complicated result, so we shall pursue a different approach that yields a simpler, albeit approximate, result that is likely to be a slight overestimate of the true (uncorrelated) error in $Q_{\rm S}(\tau)$.

The starting point is eq S27, which we now write as

$$Q_{\rm S}(\tau) \;=\; 1 \;-\; \frac{1}{\tau_{\rm R}} \int_0^\tau {\rm d}\tau'\, Q_{\rm R}(\tau') \tag{S72}$$

The discrete version of this expression is

$$Q_{\rm S}(n) \;=\; 1 \;-\; \frac{\Delta\tau}{\tau_{\rm R}} \sum_{p=0}^{n-1} Q_{\rm R}(p) \;=\; 1 \;-\; \frac{\Delta\tau}{\tau_{\rm R}} \left[ \sum_{p=0}^{n} Q_{\rm R}(p) \;-\; Q_{\rm R}(n) \right] \tag{S73}$$

or, in terms of dimensionless trajectory averages,

$$\overline{g}(n) \;=\; 1 \;-\; \frac{1}{\overline{n}} \left[ \sum_{p=0}^{n} \overline{h}(p) \;-\; \overline{h}(n) \right] \tag{S74}$$

In view of eq S59, this can be expressed as

$$\overline{g}(n) \;=\; 1 \;-\; \frac{1}{N_{\rm R}} \sum_{\alpha=1}^{N_{\rm R}} G_\alpha(n) \tag{S75}$$

where we have defined

$$G_\alpha(n) \;\equiv\; \frac{1}{\overline{n}} \left[ \sum_{p=0}^{n} h_\alpha(p) \;-\; h_\alpha(n) \right] \tag{S76}$$

We now introduce the approximation, which consists in neglecting the fluctuations in $\overline{n}$ so eq S76 becomes a linear combination of random variables,

$$G_\alpha(n) \;\approx\; \frac{1}{\langle n \rangle} \left[ \sum_{p=0}^{n} h_\alpha(p) \;-\; h_\alpha(n) \right] \tag{S77}$$

As before, we shall only consider the uncorrelated case, where $\langle n_\alpha \, n_\beta \rangle = \langle n_\alpha \rangle \langle n_\beta \rangle$ and therefore $\langle h_\alpha(n) \, h_\beta(p) \rangle = \langle h_\alpha(n) \rangle \, \langle h_\beta(p) \rangle$ and $\langle G_\alpha(n) \, G_\beta(p) \rangle = \langle G_\alpha(n) \rangle \, \langle G_\beta(p) \rangle$. It then follows immediately that

$$\sigma_0^2(\overline{g}) \;=\; \frac{\sigma_G^2}{N_{\mathrm{R}}} \tag{S78}$$

where $\sigma_G^2$ is the variance of $G_\alpha(n)$,

$$\sigma_G^2 \;=\; \left\langle [G_\alpha(n)]^2 \right\rangle \;-\; \langle G_\alpha(n) \rangle^2 \tag{S79}$$

Ensemble averaging eq S77, we obtain

$$\langle G_\alpha(n) \rangle \;=\; \frac{1}{\langle n \rangle} \left[ \sum_{p=0}^{n} \langle h(p) \rangle \;-\; \langle h(n) \rangle \right] \tag{S80}$$

Similarly,

$$\begin{aligned} \left\langle [G_\alpha(n)]^2 \right\rangle \;&=\; \frac{1}{\langle n \rangle^2} \left[ \sum_{p=0}^{n} \sum_{q=0}^{n} \langle h_\alpha(p) \, h_\alpha(q) \rangle \;-\; 2 \sum_{p=0}^{n} \langle h_\alpha(n) \, h_\alpha(p) \rangle \;+\; \left\langle h_\alpha^2(n) \right\rangle \right] \\ &=\; \frac{1}{\langle n \rangle^2} \left[ \sum_{p=0}^{n} (2p+1) \, \langle h(p) \rangle \;-\; (2n+1) \, \langle h(n) \rangle \right] \end{aligned} \tag{S81}$$

where we have made use of the identities

$$h_\alpha^2(n) \;=\; h_\alpha(n) \tag{S82}$$

$$h_\alpha(n)\, h_\alpha(p) \;=\; h_\alpha(n) \quad \text{if } p \leq n \tag{S83}$$

Combining eqs S78 – S81, we obtain

$$\sigma_0^2(\overline{g}) = \frac{1}{\langle n \rangle^2 \, N_{\mathrm{R}}} \left\{ \sum_{p=0}^{n} (2p+1)\, \langle h(p) \rangle - (2n+1)\, \langle h(n) \rangle \;-\; \left[ \sum_{p=0}^{n} \langle h(p) \rangle - \langle h(n) \rangle \right]^2 \right\} \tag{S84}$$

The estimator for the standard deviation of $\overline{g}(n)$ is thus

$$s_0(\overline{g}) \;=\; \frac{1}{\overline{n}\, N_{\mathrm{R}}^{1/2}} \left\{ \sum_{p=0}^{n} (2p+1)\, \overline{h}(p) - (2n+1)\, \overline{h}(n) \;-\; \left[ \sum_{p=0}^{n} \overline{h}(p) - \overline{h}(n) \right]^2 \right\}^{1/2} \tag{S85}$$

As for $\overline{h}(n)$, eq S85 shows that the error in $\overline{g}(n)$ has the expected property of vanishing for $n = 0$.

# References

[1] Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461 – 466.